SOK-1005 Data for Project Assignment Spring 2025

Data Acquisition

Description of the Dominick's Data is available at:

https://www.chicagobooth.edu/research/kilts/datasets/dominicks

The data manual describing the data files is available at:

https://www.chicagobooth.edu/-/media/enterprise/centers/kilts/datasets/dominicks-dataset/dominicks-manual-and-codebook_kiltscenter.aspx

Note that it is more than 500 pages, and you only need the first 28 pages if you would like to print it.

In this project, you will work on the following two datasets from the manual:

1. The customer count file. The customer count file includes information about in-store traffic. The data is store specific and on a daily basis. Use the Stata (ccount(stata).zip) format.
2. The store-specific demographic data. The data originally comes from US government (1990) census data for the Chicago metropolitan area demographics, hence there is one row of data per store. Use the Stata (demo(stata).zip) format.

Download both the customer count and store-specific demographic data. Use R or Python to compile the data for this project.

The customer count file contains total sales and total coupons redeemed (all in US dollars) for various commodity categories such as bakery, beer, bottles, etc., covering the period from 1989 to 1997. For simplicity, in this project, aggregate all the sales data (including redeemed coupons) by store and daily/monthly/year, and work with these aggregated sales data.